How to train your neural ODE

Year: 2020 | Citations: 324 | Authors: Chris Finlay, J. Jacobsen, L. Nurbekyan, Adam M. Oberman

---

Abstract

Training neural ODEs on large datasets has not been tractable due to the necessity of allowing the adaptive numerical ODE solver to refine its step size to very small values. In practice this leads to dynamics equivalent to many hundreds or even thousands of layers. In this paper, we overcome this apparent difficulty by introducing a theoretically-grounded combination of both optimal transport and stability regularizations which encourage neural ODEs to prefer simpler dynamics out of all the dynamics that solve a problem well. Simpler dynamics lead to faster convergence and to fewer discretizations of the solver, considerably decreasing wall-clock time without loss in performance. Our approach allows us to train neural ODE based generative models to the same performance as the unregularized dynamics in just over a day on one GPU, whereas unregularized dynamics can take up to 4-6 days of training time on multiple GPUs. This brings neural ODEs significantly closer to practical relevance in large-scale applications.